Unsupervised White Matter Fiber Clustering and Tract Probability Map Generation: Applications of a Gaussian Process framework for White Matter Fibers

Demian Wassermann, Luke Bloy, Efstathios Kanterakis, Ragini Verma, Rachid Deriche

INSTITUT NATIONAL DE RECHERCHE EN INFORMATIQUE ET EN AUTOMATIQUE

Unsupervised White Matter Fiber Clustering and Tract Probability Map Generation: Applications of a Gaussian Process framework for White Matter Fibers

Demian Wassermann, Luke Bloy, Efstathios Kanterakis, Ragini Verma and Rachid Deriche

N° 7005

July 2009

Thème BIO

ISSN 0249-6399 ISRN INRIA/RR--7005--FR+ENG

Unsupervised White Matter Fiber Clustering and Tract Probability Map Generation: Applications of a Gaussian Process framework for White Matter Fibers

Demian Wassermann, Luke Bloy, Efstathios Kanterakis, Ragini Verma and Rachid Deriche

Thème BIO — Systèmes biologiques
Projets Odyssée

Rapport de recherche n° 7005 — July 2009 — 32 pages

Abstract: With the increasing importance of fiber tracking in diffusion tensor images for clinical needs, there has been a growing demand for an objective mathematical framework to perform quantitative analysis of white matter fiber bundles incorporating their underlying physical significance. This paper presents such a novel mathematical framework that facilitates mathematical operations between tracts using an inner product based on Gaussian processes, between fibers which span a metric space. This metric facilitates combination of fiber tracts, rendering operations like tract membership to a bundle or bundle similarity simple. Based on this framework, we have designed an automated unsupervised atlas-based clustering method that does not require manual initialization nor an *a priori* knowledge of the number of clusters. Quantitative analysis can now be performed on the clustered tract volumes across subjects thereby avoiding the need for point parametrization of these fibers, or the use of medial or envelope representations as in previous work. Experiments on synthetic data demonstrate the mathematical operations. Subsequently, the applicability of the unsupervised clustering framework has been demonstrated on a 21 subject dataset.

Key-words: Diffusion MRI, White Matter Fiber Tracts, Gaussian Process, Tract Probability Maps, Clustering

Unité de recherche INRIA Sophia Antipolis
2004, route des Lucioles, BP 93, 06902 Sophia Antipolis Cedex (France)
Téléphone : +33 4 92 38 77 77 — Télécopie : +33 4 92 38 77 65

Regroupement et Carte de Probabilités des Faisceaux de Fibres de la Matière Blanche : Application du concept de Processus Gaussien à la description des Fibres

Résumé : Au vu de l'intérêt clinique croissant pour la tractographie des fibres en imagerie du tenseur de diffusion, la demande d'un cadre mathématique objectif pour l'analyse quantitative des fibres de la matière blanche, incorporant leur signification physique sous-jacente, s'est faite de plus en plus pressante. Cet article présente un tel cadre, facilitant les opérations mathématiques entre les faisceaux grace à un produit interne (basé sur la notion de processus gaussien) entre les fibres qui définissent alors un espace métrique. Cette métrique facilite la combinaison de faisceaux de fibres, rendant les calculs d'appartenance à un faisceau ou de similarités entre faisceaux simples. À partir de ces notions, nous avons défini une méthode automatique et non-supervisée de regroupement de faisceaux qui se base sur un atlas, et qui ne nécessite ni initialisation manuelle, ni connaissance a priori du nombre de faisceaux. On peut dès lors faire une analyse quantitative inter-sujets sur le volume des faisceaux regroupés, en évitant ainsi de devoir paramétriser les fibres, de recourrir à l'enveloppe du faisceau, ou à la fibre moyenne, contrairement aux études précédentes. Les opérations mathématiques sont montrées sur des données synthétiques, puis suit une étude de l'applicabilité de cette méthode de regroupement non supervisé, sur un jeu de 21 sujets.

Mots-clés : IRM de Diffusion, Imagerie par Tenseur de Diffusion (DTI), Procesus Gaussien, Tractographie, Regroupement, Carte de Probabilités des Faisceaux de Fibres de la Matière Blanche

Contents

1 Introduction

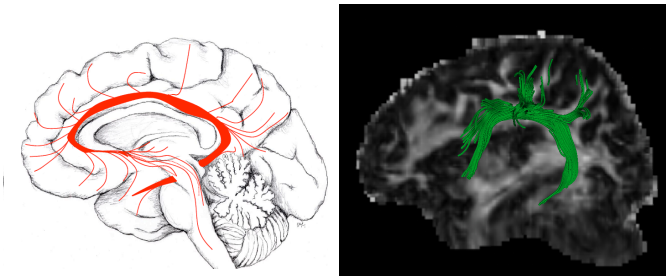
Diffusion MRI non invasively recovers the *in vivo* effective diffusion of water molecules in biological tissues. This information characterizes tissue micro-structure and its architectural organization (Basser and Pierpaoli, 1996) by modeling the local anisotropy of the diffusion process of water molecules, providing unique biologically and clinically relevant information not available from other imaging modalities. Once the diffusion information has been recovered within each voxel, it can be synthesized in the form of a diffusion tensor (Basser et al., 1994). Brain connectivity can then be assessed by assembling the tensors into tracts using tractography methods (Mori et al., 1999; Koch et al., 2002; Descoteaux et al., 2009). Among these methods, streamline tractography (Mori et al., 1999) recovers white matter fiber tracts from a seed voxel by following the principal direction of the diffusion tensor. Using this technique, white matter fiber tracts are represented as points sampled from a three-dimensional curve. Finally, these fibers can then be grouped into fiber bundles based on anatomic knowledge (O'Donnell and Westin, 2007; Maddah et al., 2008a). The Cortico Spinal Tract (CST) or the Corpus Callosum (CC) are prominent examples of the latter.

In this paper, we address the important problem of developing a mathematical framework for the quantitative analysis of fiber bundles, which has become a very active research area, with the aim of facilitating subsequent clustering and group-based statistical analysis on the bundles. The effects of pathology on tracts is evident, as in the case of brain tumors that grossly displace tracts, or can be subtle, as in the case of neuropsychiatric disorders, such as schizophrenia, which can manifest as more diffuse abnormalities in connectivity (Kubicki et al., 2007; Ciccarelli et al., 2008). We propose a framework for tract analysis based on a novel metric between bundles that serves as a probabilistic measure of inclusion of a fiber tract into a bundle. This framework is then used to develop a method for automated clustering of bundles, which are now quantifiable on the basis of membership and ready for statistics based on the distances between bundles. Once obtained, the clusters can be mapped to tract probability maps (Hua et al., 2008) enabling tract-based statistics on the cerebral white matter.

2 Materials and Methods

2.1 Prior work

The clustering of different fiber tracts into an anatomically coherent bundle, like the CC or the CST, is a challenging task for several reasons. In the first place, as seen in fig. 1, axons composing a bundle can diverge from it connecting cortical and subcortical areas. This renders approaches that quantify similarity among white matter fibers using the whole fibers instead of analyzing partial overlaps like shape statistics or rigid transformations (Veltkamp, 2001) unsuited for the clustering task. Take for instance the cingulum bundle, whose constituent fibers only partially overlap among themselves, with many diverging to innervate the cortex, as shown in fig. 1. These divergent fibers can have quite different shapes, calling



(a) Cingulum bundle (b) Arcuate fasciculus

Figure 1: Axons enter and leave anatomical bundles. Image (a) reproduced from O'Donnell (2006) with permission of Jimmy Fallon, UCI. Image (b) manually selected Arcuate fasciculus after full brain tractography.

into question the utility of shape-based metrics for subsequent tensor statistics. Moreover, even if current streamline tractography techniques are reproducible (Wakana et al., 2007), fiber tracts obtained through tractography do not recover the whole underlying axonal trajectory (Lenglet et al., 2008). This behaviour can be clearly observed in complex bundle configurations like crossings or fannings (Savadjiev et al., 2008). An example of this is the crossing area of the CC and the CST. Thus, despite innovative techniques that are more sensitive to fiber tracts being used in order to model the underlying water diffusion in a voxel and produce better local approximations to axonal distribution (Qazi et al., 2008; Descoteaux et al., 2009; Wiegell et al., 2000), results are far from being reliable after the streamline tracking procedure has traversed a region with complex bundle configuration. In fiber bundle identification (clustering tracts into a bundle), a common workaround for this problem consists of performing a dense whole brain tractography by densely seeding all over the brain. This procedure produces fiber tract sections that later could be grouped by clustering techniques (O'Donnell and Westin, 2007). This calls for the need for a similarity metric, that can quantify the closeness of two fibers or the degree to which a tract belongs to a bundle, which are both challenging problems and critical in automatic bundle integration.

Quantifying fiber similarity is a fundamental part of fiber clustering that has been addressed in different ways. Recent works (Batchelor et al., 2006; Corouge et al., 2006; Leemans et al., 2006) quantify fiber similarity with different flavors of shape statistics. However, partial overlapping of fibers is not taken into account as a similarity feature. Thus, the previous approaches are unsuited for automatic classification of fibers in the brain. There is a separate set of works (Ding et al., 2003; O'Donnell and Westin, 2007; Wassermann and Deriche, 2008; Maddah et al., 2008a), which uses different clustering algorithms based on the Hausdorff or Chamfer distances among the sequence of points parametrizing each fiber tract. This family of similarity metrics deals with sets of points instead of curves, hence they discard continuity or directionality information. Moreover, similarity tends to decrease very fast in

cases of partial overlapping, failing to include fibers diverging from the bundle in the correspondent cluster. In particular, Ding et al. (2003) only analyzes fibers whose seed points are spatially close together. This is not suited for a whole brain analysis. In this case different fiber seed points from the same bundle might have been scattered all over the white matter. The latter technique is frequently used in order to overcome limitations of streamline tracking in regions with complex bundle configurations (O'Donnell and Westin, 2007). Manifold learning techniques are used by O'Donnell and Westin (2007) and Wassermann and Deriche (2008) to generalize these type of distances from small sets of similar fibers to a bigger more diverse set of fibers. This approach embeds the fibers into Euclidean or topological spaces that can be handled more easily. This strategy has proved to be useful for single individuals but lacks the necessary parameter stability needed for group studies. O'Donnell and Westin (2007) start by generating an atlas, fibers from new subjects are then classified according to this atlas. Even though these automatically grouped bundles are anatomically coherent, the process to generate the atlas requires heavy user interaction and fine parameter tuning. The level of manual interaction needed renders the approach difficult to reproduce. Wassermann and Deriche (2008) use a publicly available anatomical atlas in conjunction with the fiber similarity metric. This work requires a smaller number of parameters, nevertheless situations of partial fiber overlapping generate non-anatomically coherent bundles. Maddah et al. (2008a) enhances the Hausdorff similarity with Mahalanobis distance between fiber points. In order to handle partial overlapping, an ad-hoc penalty term is added to this distance. This approach requires user initialization, by selecting a fiber which is known to be in the desired bundle. Their subsequent work (Maddah et al., 2008b) incorporated atlas information to increase accuracy, however an initial fiber representing each bundle is still required. From all the presented approaches, only O'Donnell and Westin (2007) succeed in the task of semi-automated classification of the whole ensemble of white matter fibers, however this is achieved with a great deal of user interaction and parameters tuning.

Once bundles have been found, quantitative analysis can be performed. This analysis is useful to monitor pathological conditions (Ciccarelli et al., 2008; Kubicki et al., 2007). Most of the works performing bundle statistics (Goodlett et al., 2008; Hua et al., 2008; O'Donnell et al., 2007; Maddah et al., 2008a) rely on the use of medial representations for bundles. These representations are only appropriate for bundles which can be modeled as convex envelopes. Thus, their methodology is not entirely appropriate at the extremes of the bundles, an area where the axons fan-out inervating cortical or subcortical structures. Other works (Oh et al., 2007) use a mesh approach over ROIs and perform statistics on the surface. However, automated transition from a set of fibers to the mesh is unclear. There is recent evidence that tract probability maps can be used in order to perform bundle-oriented statistics in diffusion MRI (Hua et al., 2008) and histological images (Bürgel et al., 2006). However, in these two approaches the process to obtain the tract probability maps must be performed manually by experts. Overall, statistical models of white matter bundles which do not rely on medial representations are scarce. Moreover, for models which are more appropriate for a wider spectra of bundles, the transition from automatically obtained bundles to these statistical models is not straightforward.

It is the goal of this work to introduce a novel mathematical framework to model the combination of fiber tracts and bundles. Our model includes diffusion information and relates the bundles with a ROI in the volume, mapping every voxel to its probability of belonging to the bundles, *the tract probability map*. Our framework provides an inner product space for fiber bundles and fiber tracts, which are considered as single-fiber bundles. Inherently, it provides a metric between these elements. Moreover, linear combinations of fiber bundles seamlessly generate new bundle configurations and allow for the volume-based statistics of fiber bundles. In addition, the framework handles cases of partial fiber overlap, in fiber combinations, naturally. Furthermore, the previously mentioned characteristics of our framework facilitate statistical analysis and classification/clustering tasks. Finally, we present a clustering application based on our mathematical framework and on anatomical information in the shape of a volumetric atlas. The output of this application is a set of automatically obtained white matter bundles like the Arcuate Fasciculus or the Cingulum. For each bundle, we are also able produce a ROI which maps every voxel to its probability of belonging to the bundle, referred in previous work as *tract probability map* (Hua et al., 2008; Bürgel et al., 2006). This map is an appropriate tool to perform bundle-based statistics on the cerebral white matter. We validate this clustering algorithm by applying it to 21 healthy subjects and then performing statistical analysis on the results using our mathematical framework.

2.2 A Mathematical framework for white matter fibers and bundles

In this section, we develop our mathematical framework for the quantitative analysis of fiber bundles with the aim of facilitating subsequent clustering and group-based statistical analysis on the bundles. Within this framework, examples of operations such as the combination of fibers into bundles, similarity measurements between fibers, as well as *tract probability maps* computation will be easily provided. To this end, we first need to model each fiber as a stochastic process. This is done using the well known Gaussian Process framework (MacKay, 1998) which has now a long history in the field of statistics. Indeed, a Gaussian process can be seen as a generalization of the classical Gaussian probability distribution to describe properties of functions and not only properties of random variables such as scalars or vectors. A stochastic process will be defined to model each fiber as a function. This will allow us to govern the properties of our fibers/functions while subsequently simplifying a certain number of sophisticated operations between the fibers. Examples of linear combination of Gaussian Processes of several fibers resulting in the Gaussian Process of a bundle are shown in fig. 3. It is important to point out that such framework provides also an adequate computational tractability. It allows us to measure bundle similarity with partial fiber overlap without relying on point correspondences as shown in fig. 4. Furthermore, the linear combination and inner product operations span an inner product space, which is an appropriate framework to perform statistical analyses such as clustering on fiber bundles. Finally, as we model bundles with stochastic processes, producing a *tract probability map* for each bundle is a natural operation. Thus, the mathematical framework we develop in this section provides all the right and necessary tools to perform a quantitative statistical

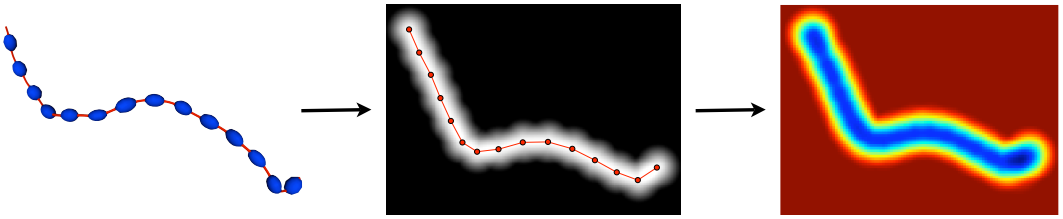


Figure 2: Gaussian Process representation of the fiber: We model each fiber as a blurred indicator function. Each fiber obtained by tractography is constituted as a sample of points over a line, which is shown in red line on the left side of the figure, and its corresponding diffusion tensor field, indicated by the blue tensors over the red fiber. The sampled points are then used to generate a blurred indicator function with a maximal level set which corresponds to the fiber (central image of the figure). Then, the fiber is blurred in accordance with the information provided by the underlying diffusion tensor field. The resulting representation, on the right side of the figure, is a smooth indicator function blurred along the fiber direction.

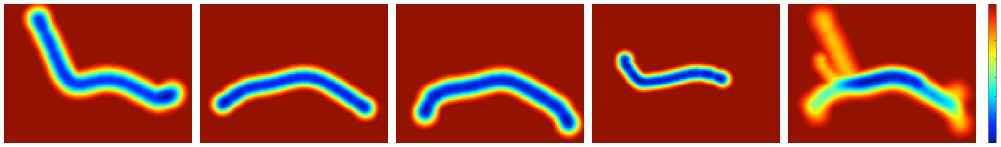
analysis of white matter fiber bundles. We model each fiber tract $\mathcal{F}$ as a *blurred indicator function* $y_{\mathcal{F}} : \mathbf{p} \in \mathbb{R}^3 \rightarrow \mathbb{R}$. This *blurred indicator function* has a maximal level set which corresponds to $\mathcal{F}$. Moreover, $y_{\mathcal{F}}$ is blurred in accordance with the information provided by the underlying diffusion tensor field, i.e. along the fiber direction and not across it. This is illustrated on fig. 2. As we show in appendix A, Gaussian Processes (GP) provide the right framework to integrate spatial and diffusion information for $y_{\mathcal{F}}$. More precisely, the *blurred indicator function* $y_{\mathcal{F}}(\mathbf{p})$ is modeled by the GP

$$y_{\mathcal{F}}(\mathbf{p}) \sim \mathcal{GP}(y^*_{\mathcal{F}}(\mathbf{p}), c_{\mathcal{F}}(\mathbf{p}, \mathbf{p}')), \tag{1}$$

where the mean function $y^*_{\mathcal{F}}(\mathbf{p})$ and covariance function $c_{\mathcal{F}}(\mathbf{p}, \mathbf{p}')$ are the parameters of this stochastic process. These two functions are inferred from the tractography of each fiber. That is, from the sequence of points $\mathbf{f} = \{\mathbf{f}_1, \ldots, \mathbf{f}_{|\mathbf{f}|}\}$ estimated by tractography of the anatomical bundle $\mathcal{F}$ and from the corresponding sampling on its tensor field $\Sigma(\mathbf{f}_1), \ldots, \Sigma(\mathbf{f}_{|\mathbf{f}|})$. The inference process of the GP corresponding to a white matter fiber is developed in appendix A.1. Using our framework, we are able to combine N fibers into a bundle with a simple average operation in our inner product space. The GP which corresponds to the indicator function of a fiber bundle $\mathcal{B}$, is obtained through the mean Gaussian Processes of single fibers or smaller bundles composing it,

$$y_{\mathcal{B}}(\mathbf{p}) = \frac{1}{N} \sum_{i=1}^{N} y_{\mathcal{F}_i}(\mathbf{p}) \sim \mathcal{GP}\left(y^*_{\mathcal{B}}(\mathbf{p}); c_{\mathcal{B}}(\mathbf{p}, \mathbf{p}')\right), \tag{2}$$

$$\text{where } y^*_{\mathcal{B}}(\mathbf{p}) = \frac{1}{N} \sum_{i=1}^{N} y^*_{\mathcal{F}_i}(\mathbf{p}) \text{ and } c_{\mathcal{B}}(\mathbf{p}, \mathbf{p}') = \frac{1}{N^2} \sum_{i=1}^{N} c_{\mathcal{F}_i}(\mathbf{p}, \mathbf{p}').$$



(a) $\mathcal{F}_1$ (b) $\mathcal{F}_2$ (c) $\mathcal{F}_3$ (d) $\mathcal{F}_4$ (e) $\sum_i \frac{\mathcal{F}_i}{4}$

Figure 3: Mean indicator function for four fiber tracts (a-d) and mean indicator function for the bundle formed by averaging them according to our framework (e). Blue color means that the bundle is more likely to cross that voxel while red color means it is not likely that the bundle traverses that voxel. Fibers were manually selected from a full brain tractography and belong to the Cingulate Cortex section of the Cingulum (CgC).

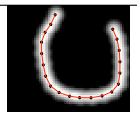
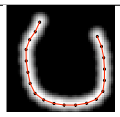
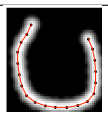
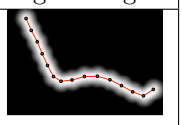
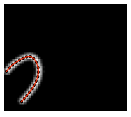
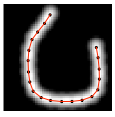
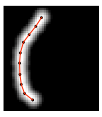
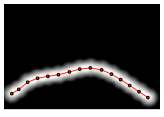
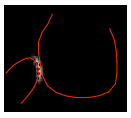
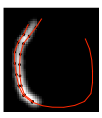
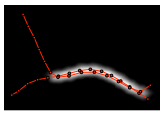
	FF & UNC	FF & FF	FF & FF	CgC & CgC
$y^*_{\mathcal{F}}(\mathbf{p})$				
$y^*_{\mathcal{F}'}(\mathbf{p})$				
$y^*_{\mathcal{F}}(\mathbf{p})y^*_{\mathcal{F}'}(\mathbf{p})$				
$\langle \mathcal{F}, \mathcal{F}' \rangle$	4.26	115.61	48.03	2089.
$\frac{\langle \mathcal{F}, \mathcal{F}' \rangle}{\|\mathcal{F}\|\|\mathcal{F}'\|}$	0.093	0.800	0.474	0.600

Figure 4: Examples of the product of *blurred indicator functions* for different fiber pairs, the value of our inner product operation $\langle \mathcal{F}, \mathcal{F}' \rangle$, defined in eq. 3, and of our inner product normalized by its natural norm, $\|\mathcal{F}\| = \sqrt{\langle \mathcal{F}, \mathcal{F} \rangle}$. Inner product quantifies the overlapping of *blurred indicator functions*. A larger inner product means that fibers are more similar and relates to the volume of the overlapping. The inner product normalized by its norm quantifies similarity ranging from 0 when overlapping is null to 1 when the two fibers are identical. The compared fibers have been extracted from different anatomical tracts of a diffusion MRI image, the Frontal Forceps (FF), the Uncinate Fasciculus (UNC) and the Cingulate Cortex section of the Cingulum (CgC).

An example of this is shown in fig. 3, where indicator functions for four fiber tracts and the obtained function for the bundle combining them can be seen. Within the proposed

framework, we measure similarity between bundles by quantifying their overlapping. This is done with an inner product operation developed in detail in appendix A.3. If the mean function for a bundle $\mathcal{F}$ is square integrable and has finite support, the inner product between two bundles $\mathcal{F}$ and $\mathcal{F}'$ is defined as

$$\langle \mathcal{F}, \mathcal{F}' \rangle := \int_{\mathbb{R}^3} y^*_{\mathcal{F}}(\mathbf{p}) y^*_{\mathcal{F}'}(\mathbf{p}) d\mathbf{p}, \tag{3}$$

along with its induced norm

$$\|\mathcal{F}\|^2 := \langle \mathcal{F}, \mathcal{F} \rangle. \tag{4}$$

Moreover, we can define a similarity measure, bounded by 1 when $\mathcal{F}$ and $\mathcal{F}'$ are exactly the same and by 0 when there is no intersection, as the inner product normalized by its induced norm,

$$\langle \mathcal{F}, \mathcal{F}' \rangle_N := \frac{\langle \mathcal{F}, \mathcal{F}' \rangle}{\|\mathcal{F}\| \|\mathcal{F}'\|}. \tag{5}$$

Examples of these two similarity measures are shown in fig. 4.

We mentioned previously the importance for bundle statistics of the *tract probability map*, the probability that a point $\mathbf{p}$ in $\mathbb{R}^3$ is contained in a bundle $\mathcal{B}$, $\mathbb{P}\{\mathbf{p} \in \mathcal{F}\}$. As we show in appendix A.4, having calculated the GP for a bundle $\mathcal{B}$, $y_{\mathcal{B}}(\cdot)$, by means of eq. 2, we can define the *tract probability map* from the parameters of the GP as

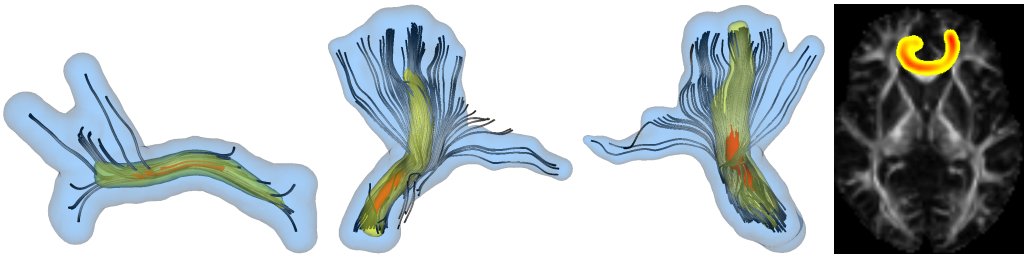
$$\mathbb{P}\{\mathbf{p} \in \mathcal{F}\} \propto \frac{1}{2\sqrt{\pi(h^2 + \sigma^2_{\mathcal{B}}(\mathbf{p})}}, \tag{6}$$

where h is a parameter that diffuses the *tract probability map* in space and $\sigma^2_{\mathcal{B}}(\mathbf{p})$ is calculated from the parameters of $y_{\mathcal{B}}(\cdot)$. In order to illustrate the *tract probability map* calculated from a GP for a bundle, color-coded surfaces and a probability map over an FA image for anatomical bundles are shown in fig. 5.

Up to this point, we have introduced our GP-based framework for white matter fiber bundles and its three main operations: combination of fibers into a bundle, similarity quantification and calculation of the *tract probability map*. In the following sections we use these tools in order to perform automatic white matter bundle identification by means of white matter fiber clustering and we assess the quality of this clustering by performing inter-subject statistical analysis of white matter fiber bundles using our framework.

2.3 Clustering Algorithm

Taking advantage of the mathematical framework for fiber bundles presented in section 2.2, we propose a stochastic process-based agglomerative clustering algorithm. This algorithm is executed over a set of densely sampled fibers from the whole white matter, a full brain tractography. Once executed, our algorithm generates a dendrogram, a tree structure where each joint is a candidate cluster, this is illustrated in fig. 6. Then, this dendrogram can be interactively explored in order to choose the desired granularity of the clustering without



(a) $\mathcal{B}_1$: right CG (b) $\mathcal{B}_2$: right CST (c) $\mathcal{B}_3$: left CST (d) $\mathcal{B}_4 : FF$

Figure 5: Iso-probability surfaces for manually selected fiber bundles, fig. (a-c), and tract probability map over FA for an automatically obtained bundle, fig. (d), see section 2.3. The probability at each voxel is calculated using eq. 6. Color code for fig. (a-c) is as follows, Blue: $\mathbb{P}\{\mathbf{p}\in\mathcal{B}_i\} = .01$, Yellow: $\mathbb{P}\{\mathbf{p}\in\mathcal{B}_i\} = .2$, Red: $\mathbb{P}\{\mathbf{p}\in\mathcal{B}_i\} = .6$.

reprocessing the data. Furthermore, being a hierarchical agglomerative algorithm it has several desirable properties. To begin with, convergence is guaranteed by the finite number of elements to cluster. Next, the number of clusters does not have to be known *a priori*. Furthermore, outliers are incorporated to clusters during the late stages of the process if at all, thus being easily distinguishable from clustered bundles (Jain et al., 1999). All of these characteristics make our clustering algorithm effective and robust in order to classify white matter fibers from a full brain tractography into anatomically coherent bundles.

Our clustering algorithm applied to a full brain tractography is as follows,

Step 1. Given a full brain tractography $F = \{\mathcal{F}_i\}$, with $1 \leq i \leq |F|$, calculate the set of stochastic processes representing each fiber $Y = \{y_{\mathcal{F}_i}(\mathbf{p})\}$.

Step 2. Initialize the clustering as the set of single fiber bundles $B = \{\mathcal{B}_i\}$, where $\mathcal{B}_i$ is a set (bundle) formed by a single fiber, $\mathcal{F}_i$, and $1 \leq i \leq |F|$.

Step 3. Initialize the set of edges in the dendrogram: $T = \emptyset$

Step 4. While there is a pair of different bundles $\mathcal{B}$,$\mathcal{B}'$ in B, s.t. $\langle \mathcal{B}, \mathcal{B}' \rangle > 0$

  Step 4.1. Select two different bundles $\mathcal{B}, \mathcal{B}'$ such that $\langle \mathcal{B}, \mathcal{B}' \rangle = \max_{\mathcal{C},\mathcal{C}'\in B} \langle \mathcal{C}, \mathcal{C}' \rangle$.

  Step 4.2. Remove the bundles $\mathcal{B}$ and $\mathcal{B}'$ from B and add the bundle $\{\mathcal{B} + \mathcal{B}'\}$.

  Step 4.3. Add the edge $(\mathcal{B}, \mathcal{B}')$ to the dendrogram T.

The output of this algorithm is a dendrogram T, more precisely set of trees where each edge represents the joining of two bundles such as the one shown in fig. 6.

A main advantage of our framework within this clustering algorithm is that the most important operation in step 4, the inner product among bundles described in section 2.2,

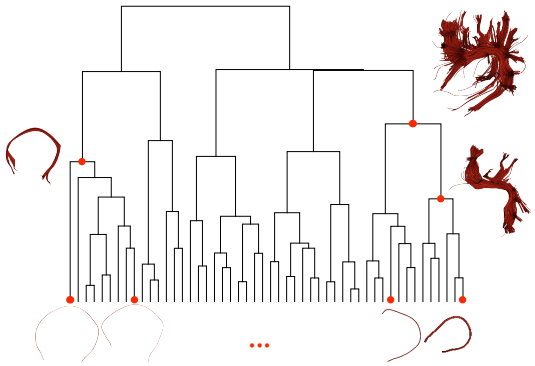


Figure 6: Illustration of our clustering algorithm defined in section 2.3. The algorithm starts from all the single-fiber bundles obtained by a full brain tractography, shown at the bottom of the figure, and joins them into multi-fiber bundles according to our similarity measure (eq. 3). This generates a tree structure called dendrogram. Finally every joint is a candidate cluster. Sample clusters are shown on the side of the dendrogram and their positions on the dendrogram are marked with red dots.

is fast and simple to compute as we show in appendix A.4. Thus, once we have calculated the matrix of inner products for every pair of fibers in the full brain tractography $\mathcal{F}$, the algorithm works by simply performing linear operations on the rows of this matrix.

2.4 Tract Querying: Automatic cluster selection based on anatomical knowledge

Once our clustering algorithm generates the dendrogram for a full brain tractography, section 2.3, the main problem is how to select the joints in the dendrogram so that they are anatomically correct clusters. In order to do this, we introduce a query system based on volumetric information, the result of this query will be a cluster selected from joints of the dendrogram. In this work we use a publicly available atlas which has a parcellation of the brain gyri on the grey and white matter (Wakana et al., 2004) as anatomically-aware volumetric information. Then, an anatomical query is defined by a set of grey or white matter regions that the tract must traverse, for example, the *Inferio Fronto Occipital* tract must connect the *inferio-frontal gyrus* and the *medial-occipital gyrus*. Images of the white matter atlas parcellation and queries for various tracts are shown in fig. 7. After setting an anatomical query Q traversing K labeled regions on the atlas, $Q = \{r_1, \ldots, r_K\}$, we use the *tract probability map* of each bundle, eq. 6, to select the bundle $\mathcal{B}$ on the dendrogram T with

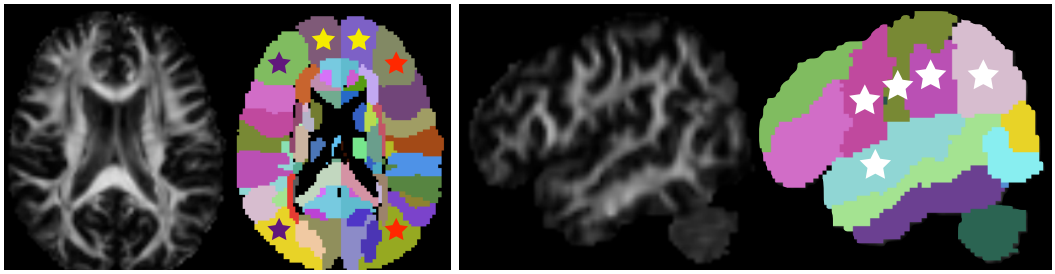


Figure 7: Illustration of anatomical queries based on volumetric information. A Fractional Anisotropy image and an image of a parcellation of the white matter are shown. The colored stars indicate the anatomical queries. On the left: Purple and red stars tag the inferio-frontal gyrus and middle-occipital gyrus on the left and right sides respectively, this query corresponds to the inferio-fronto-occipital fasciculus on the left and right hemispheres. Yellow stars tag the left and right middle-frontal-orbital gyrus, this query corresponds to the frontal forceps. On the right: white stars tag the pre and post central gyri, the angular, supra-marginal and superio-temporal gyri, this query corresponds to the arcuate fasciculus. The results of this queries on 4 subjects are exhibited in fig.s 9 to 12 and *tract probability maps* for the mean bundles for every subject are shown in fig.s 13 and 14.

maximal joint probability of traversing all the regions:

$$\mathcal{B} = \underset{\mathcal{B}' \in T}{\operatorname{argmax}} \prod_{r \in Q} \int_{\mathbf{p} \in r} \mathbb{P}\left\{\mathbf{p} \in \mathcal{B}'\right\} d\mathbf{p}. \tag{7}$$

The results of queries for 13 different white matter tracts, table 1, on 4 subjects are exhibited in fig.s 9 to 12 and *tract probability maps* of the mean bundles across subjects are shown in fig.s 13 and 14.

3 Experiments

3.1 Subjects, Imaging and Data Processing

Whole-brain DWI datasets were acquired from 21 healthy volunteers (30.05 +- 7.05 years, 10 Male) on a Siemens Trio 3T scanner with 1.71x1.71mm^2 in-plane resolution, 2mm thick slices, six unweighted images and 64 diffusion weighted images (b=1000s/mm^2) acquired with non-collinear diffusion sensitizing gradients.

DTI images for each subject were computed and deformably registered, using DTI-DROID (Yang et al., 2008), to a DTI atlas (Wakana et al., 2004). Full brain tractography was performed following (O'Donnell and Westin, 2007), streamline tractography was performed by seeding in sub-voxel resolution by taking every voxel with linear anisotropy higher

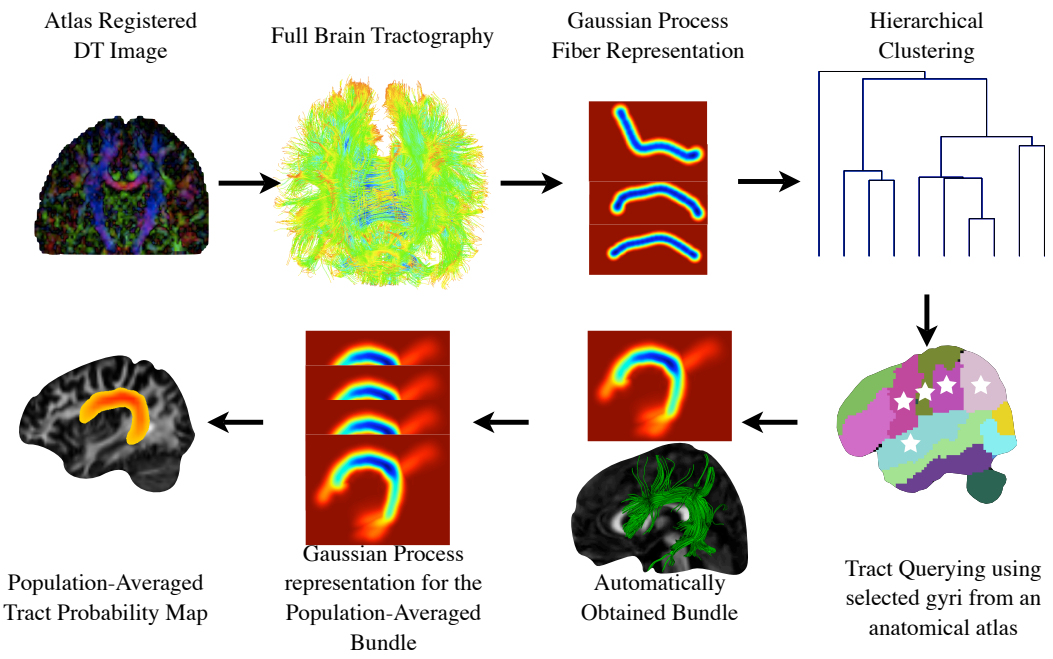


Figure 8: The procedure used to cluster white matter fibers into anatomical bundles and produce the *tract probability maps for each bundle*: We registered 21 Diffusion Tensor MR images using (Yang et al., 2008). We performed full brain tractography obtaining around 10.000 fibers per brain. We produced the Gaussian Process representation for each fiber as we describe in section 2.2. We identified anatomical bundles like the arcuate or the uncinate fasciculus by applying our clustering and tract querying algorithms to each subject individually as we describe in section 2.3. Finally we produced a population-averaged Gaussian Process for each identified bundle and the corresponding *tract probability map* with the methodology described in section 2.2

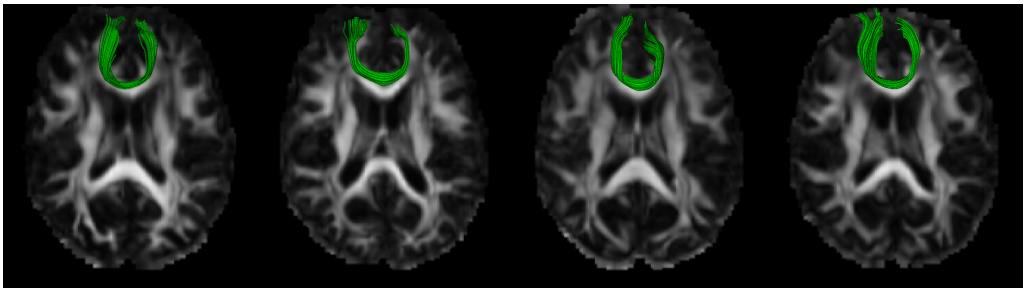
than .3, dividing it into .25x.25x.25mm^3 sub-voxels and seeding from each sub-voxel. In average, 10.000 fiber tracts where obtained for each subject. Then, the previously presented clustering algorithm based on our mathematical framework was applied to every subject individually. In order to extract major white matter tracts on every subject individually we performed set of queries, shown in table 1, over the dendrogram obtained from the clustering of each subject using the tract querying algorithm we introduced. The whole process is shown in fig. 8

White matter tract to extract	Tract queries
frontal forceps	middle frontal orbital left gyrus, middle frontal orbital right gyrus
posterior forceps	medial occipital left gyrus, medial occipital right gyrus
fornix (left-section)	fornix left, medial temporal left gyrus
cortico spinal tract (post-central)	midbrain, post central gyrus
uncinate fasciculus	middle frontal orbital gyrus, medio temporal gyrus
arcuate fasciculus	post central gyrus, pre central gyrus,angular gyrus, supra marginal gyrus, superio temporal gyrus
cingulum bundle	cingulate gyrus
inferio fronto occipital tract	medial occipital gyrus, inferio frontal gyrus

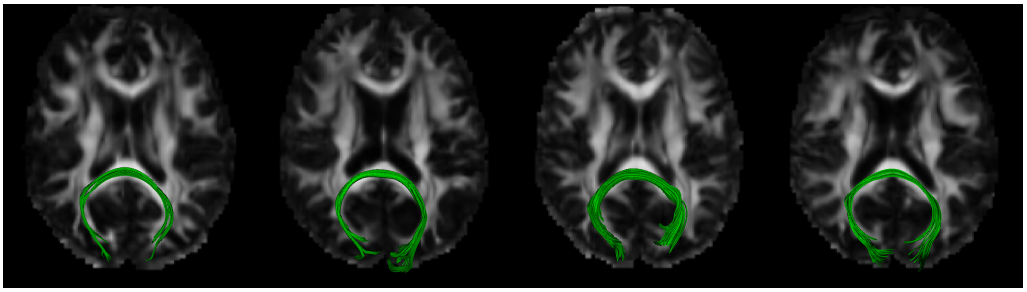
Table 1: Queries applied to a set of 21 registered subjects in order to automatically extract major white matter fiber bundles using the algorithm presented in section 2.3. The results of this queries on 4 subjects are exhibited in fig.s 9 to 12 and *tract probability maps* for the *tract probability maps* corresponding to the population-averaged bundles for every subject are shown in fig.s 13 and 14

4 Results

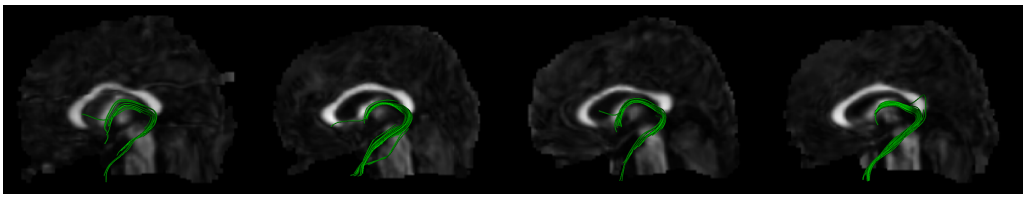
We show the results of our clustering and tract querying algorithms for 4 different subjects selected from our 21 subject database in fig.s 9 to 12. Then, we evaluated cluster coherence across subjects. Firstly, calculation of the GP representation of the population-averaged bundle across the 21 subjects for each automatically extracted tract: For each queried white matter tract, the GP corresponding to the population-averaged bundle across subjects was calculated using eq. 2. We show the *tract probability map* for each population-averaged bundle, calculated with eq. 6, in fig.s 13 and 14. These maps are in atlas space, over fractional anisotropy images. Finally, in order to provide quantitative evaluation, for each bundle extracted from each subject we quantified its similarity with respect to the corresponding population-averaged bundle using the normalized inner product of our mathematical framework, eq. 2.2. The result is plotted in fig. 15. In this plot, the boxes span between the second and third quartiles of the similarity with the population-averaged bundle, the red bar is the median similarity. Moreover, the whiskers indicate the bundles whose similarity value with the population-averaged bundle is the smallest and largest within within 1.5 times the interquartile distance of the population-averaged similarity and the '+' symbols indicate outliers.



(a) Frontal Forceps



(b) Posterior Forceps

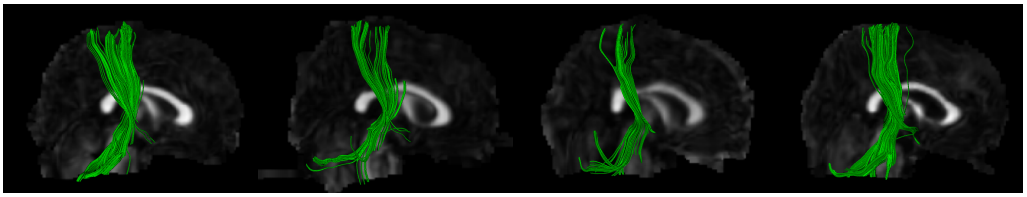


(c) Left Fornix

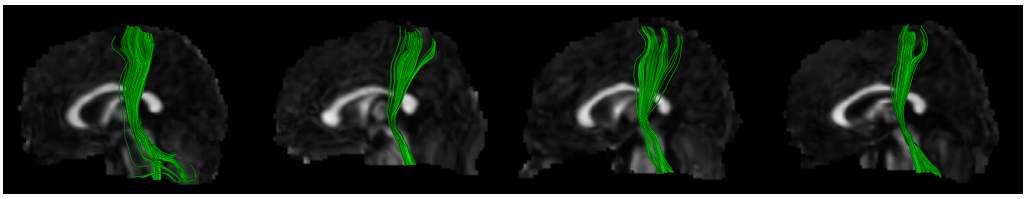
Figure 9: White matter fiber bundles automatically clustered from a database of 21 subjects using the clustering algorithm presented in section 2.3 and the queries presented on table 1. The subjects where previously registered to a white matter fiber atlas and full brain tractography was performed by densely seeding on the whole cerebral white matter. Four subjects are shown.

5 Discussion

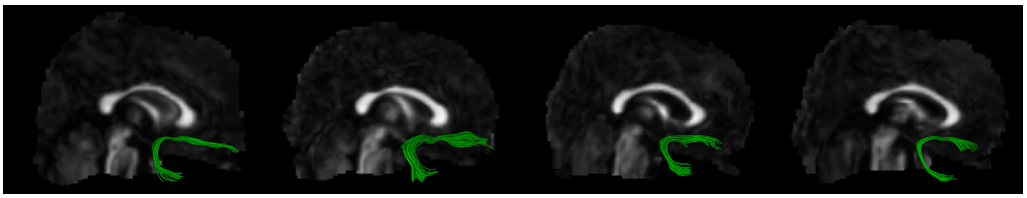
In recent years tractography has become a popular means to perform white matter studies through diffusion MRI (Ciccarelli et al., 2008). Tractography results are visually appealing and recent developments produced means to perform analysis of diffusion-derived measures



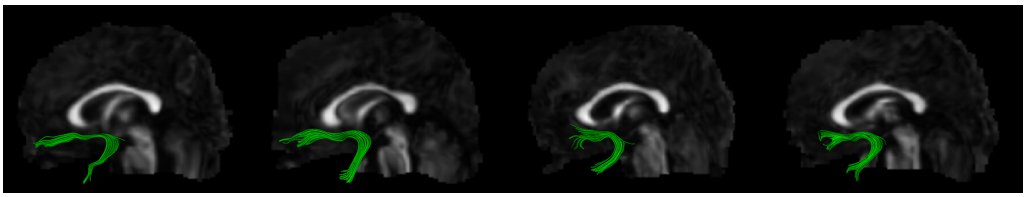
(a) Cortico Spinal Tract right



(b) Cortico Spinal Tract left



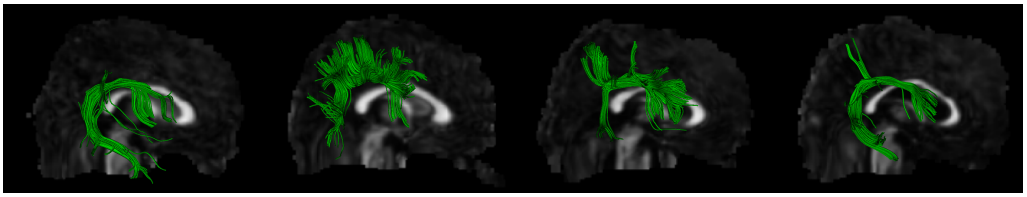
(c) Uncinate Fasciculus left



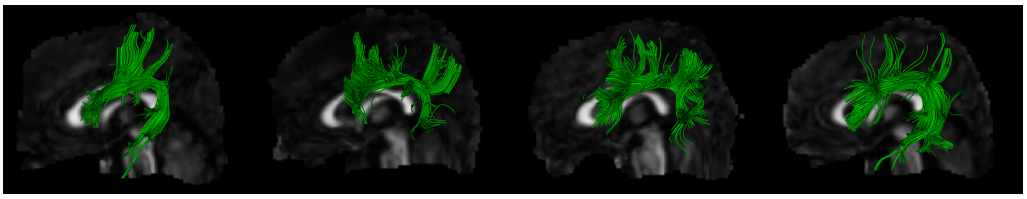
(d) Uncinate Fasciculus left

Figure 10: White matter fiber bundles automatically clustered from a database of 21 subjects using the clustering algorithm presented in section 2.3 and the queries presented on table 1. The subjects where previously registered to a white matter fiber atlas and full brain tractography was performed by densely seeding on the whole cerebral white matter. Four subjects are shown.

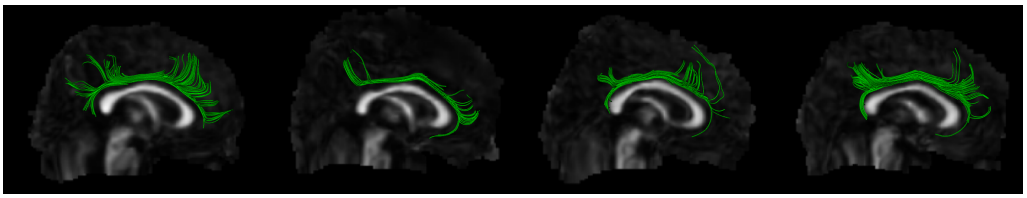
like the Fractional Anisotropy. However, statistical analysis on the fibers themselves has not been performed in sound ways due to the lack of an appropriate mathematical framework. White matter fiber clustering (O'Donnell and Westin, 2007; Wassermann and Deriche, 2008;



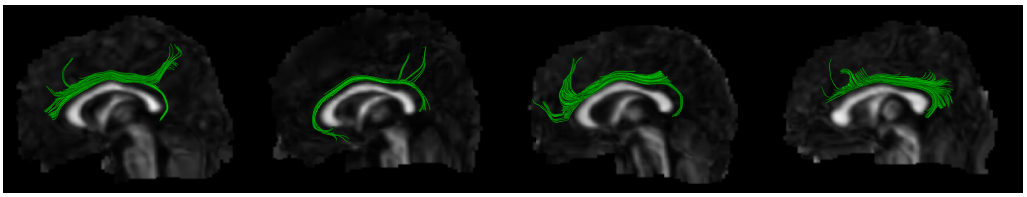
(a) Arcuate fasciculus right



(b) Arcuate fasciculus left



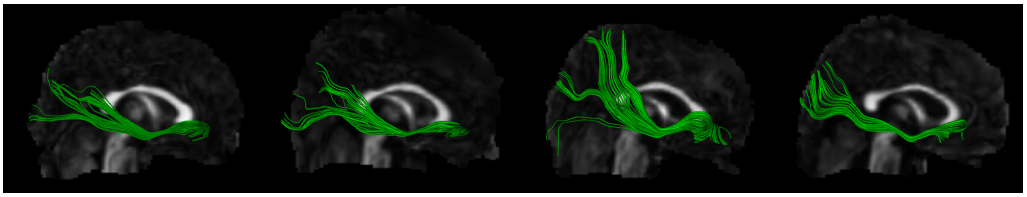
(c) Cingulum right



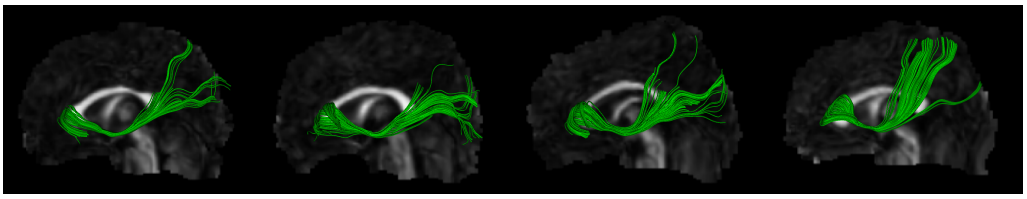
(d) Cingulum left

Figure 11: White matter fiber bundles automatically clustered from a database of 21 subjects using the clustering algorithm presented in section 2.3. The subjects where previously registered to a white matter fiber atlas and full brain tractography was performed by densely seeding on the whole cerebral white matter. Four subjects are shown.

Maddah et al., 2008a) and tract probability maps have been previously reported (Hua et al., 2008). However, clustering approaches were not suited for large-scale clinical studies or required a great deal of parameter tuning. With regard to *tract probability maps*, they were



(a) Inferio Fronto Occipital fasciculus right



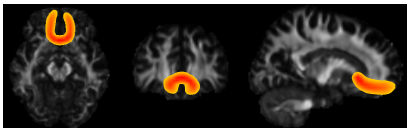
(b) Inferio Fronto Occipital fasciculus left

Figure 12: White matter fiber bundles automatically clustered from a database of 21 subjects using the clustering algorithm presented in section 2.3. The subjects where previously registered to a white matter fiber atlas and full brain tractography was performed by densely seeding on the whole cerebral white matter. Four subjects are shown.

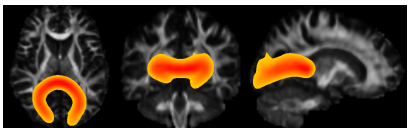
calculated by averaging binary masks obtained by manually extracted white matter fibers from tractography or by chemical staining.

This paper proposes a mathematical framework that provides the necessary tools to perform automatic clustering and identification of white matter fibers and subsequent statistics on them. Thus, automatic clustering of white matter fibers into bundles which is consistent among individuals is a fundamental tool for clinical studies. In this work we provide the tools to solve these issues. Firstly, we develop a mathematical framework for handling white matter bundles which includes spatial and diffusion information and provides the grounds for their statistical analysis. Secondly, we use this framework to develop a clustering algorithm and automatically obtain white matter structures among subjects, which is then applied to 21 subjects.

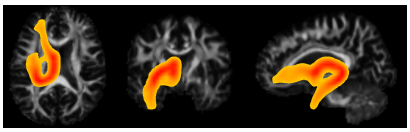
Our mathematical framework for white matter bundles, sets the grounds for statistical analysis of white matter fiber bundles including applications like clustering and tract-based quantification of scalar quantities among others. It provides three important operations for white matter bundle statistics: 1) an operation to merge white matter fibers into a bundle by means of a linear combination within our framework, see Fig. fig. 3; 2) an operation to quantify bundle similarity based on their overlap in space (see Fig. fig. 4, quantifies the volume of the overlapping region between two fibers and can be normalized in order to provide a similarity index ranging from 1, when fibers are exactly alike, to 0, when they are completely different. Furthermore, we provide the means to calculate the probability



(a) Frontal Forceps



(b) Posterior Forceps

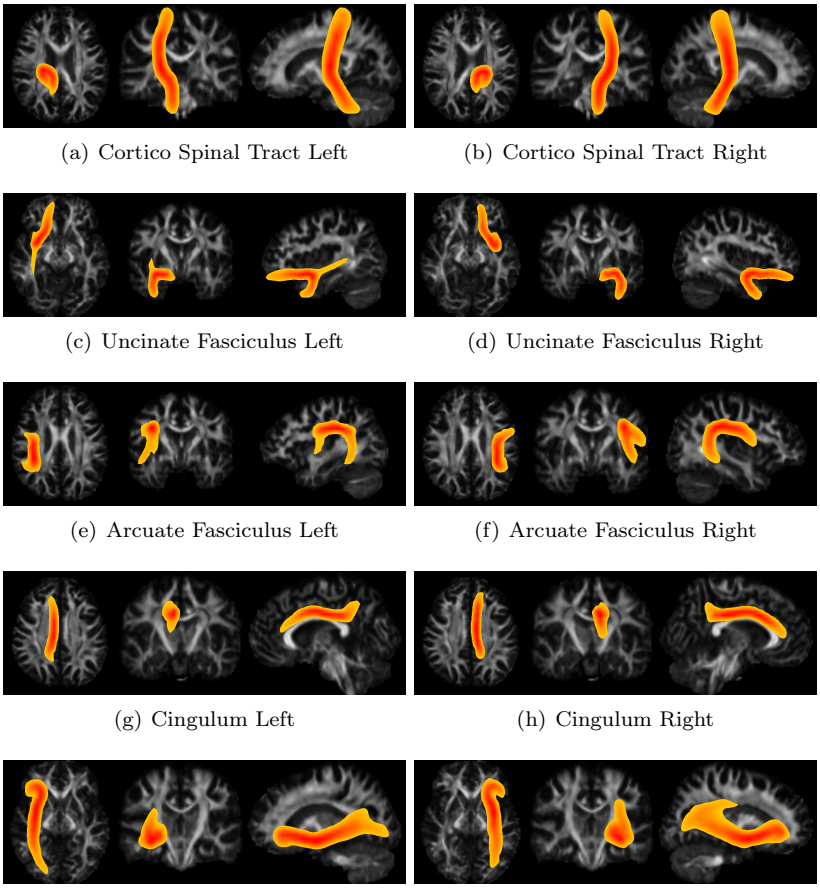


(c) Left Fornix

Figure 13: *Tract probability maps* for the averaged white matter fiber bundles over 21 subjects. The bundles have been automatically extracted from each subject individually using the clustering algorithm presented in section 2.3 and the queries presented in table 1. The subjects where previously registered to a white matter fiber atlas and full brain tractography was performed by densely seeding on the whole cerebral white matter. Maximum intensity projection is used for the color intensity. Color code ranges from red, when the probability of the voxel belonging to the bundle is 1.0 to yellow, when the probability of the voxel belonging to the bundle is 0.2.

that a point in space belongs to a bundle, the *tract probability map* of the bundle. *Tract probability maps* were recently introduced as a tool to perform bundle-specific quantification of scalar quantities like Fractional Anisotropy (Hua et al., 2008). Moreover, *tract probability maps* have a wide range of applications. For instance, in this work we use them to relate the results of our clustering algorithm with anatomical information given by a white matter atlas. This enables us to perform consistent identification of white matter structures among subjects. These maps could aid as a feature in DTI registration

As part of the clustering, we create a dendrogram, fig. 6 from the full brain tractography. The dendrogram is a versatile tool to represent clustering results, ita a precomputed tree structure where each joint is a candidate cluster. Among other advantages, it allows interactive exploration of the clustering results at different scales and it is resilient to outliers. Moreover, clusters can be selected at different scales using different postprocessing methodologies. In particular, we take advantage of this feature using the *tract probability map* for



(a) Cortico Spinal Tract Left (b) Cortico Spinal Tract Right

(c) Uncinate Fasciculus Left (d) Uncinate Fasciculus Right

(e) Arcuate Fasciculus Left (f) Arcuate Fasciculus Right

(g) Cingulum Left (h) Cingulum Right

(i) Inferio Fronto Occipital Left (j) Inferio Fronto Occipital Right

Figure 14: *Tract probability maps* for the averaged white matter fiber bundles over 21 subjects. The bundles have been automatically extracted from each subject individually using the clustering algorithm presented in section 2.3 and the queries presented in table 1. The subjects where previously registered to a white matter fiber atlas and full brain tractography was performed by densely seeding on the whole cerebral white matter. Maximum intensity projection is used for the color intensity. Color code ranges from red, when the probability of the voxel belonging to the bundle is 1.0 to yellow, when the probability of the voxel belonging to the bundle is 0.2.

each cluster and a volumetric atlas with a parcellation of the cerebral gyri. We combine both of these tools with the dendrogram in order to produce a *tract query* algorithm.

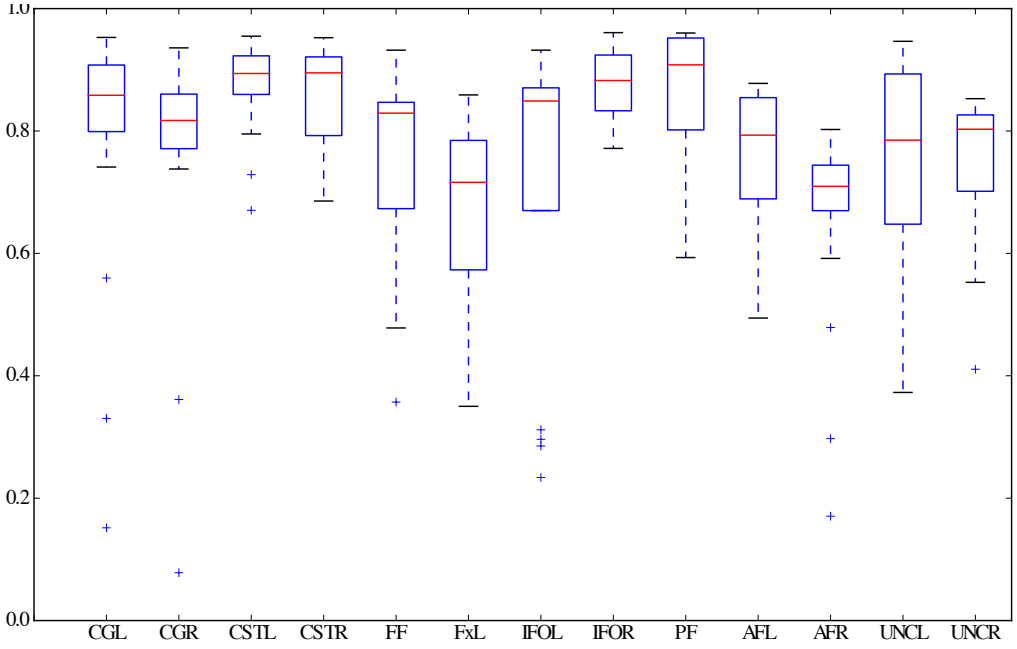


Figure 15: Quantitative assessment of bundle coherence among subjects. For each automatically extracted bundle, the similarity between the mean bundle across all subjects and the bundle extracted from each subject individually was calculated by means of the normalized inner product, eq. 2.2. Similarity ranges from 1, bundles are identical, to 0, bundles are completely different. Boxes span between the second and third quartiles of the similarity with the mean bundle, the red bar is the median similarity. Moreover, the whiskers indicate the bundles whose similarity value with the mean bundle is the smallest and largest within 1.5 times the interquartile distance of the mean similarity and the '+' symbols indicate outliers.

We apply our clustering and querying process to 21 healthy subjects using queries for 13 white matter structures shown in table 1. This produces a set of white matter fiber bundles shown in fig.s 9 to 12. Then, we use our mathematical framework to generate a population-averaged GP and *tract probability maps* for each identified bundle. These maps allow qualitative analysis. The result of the clustering on the whole population can be analyzed through visual inspection of the population-averaged *tract probability maps*, fig.s 13 and 14. They allow quantitative analysis: measuring the similarity of the bundle

corresponding to each subject against the population-averaged GP for each bundle, fig. 15, allows us to quantify and analyze the dispersion of the similarity within the population.

Next we look at the qualitative and quantitative assessment of the results of our clustering. Firstly, we note that tracts obtained by means of our clustering-querying procedures (fig.s 9 to 12), are consistent with manually obtained tracts by experts in diffusion MRI images (Wakana et al., 2004, Figures 3, 4, and 5) and macroscopical preparations (Lawes et al., 2008, Figures 5iii and 6ii). Since our similarity metric handles partial overlap of fibers, we are able to correctly cluster fibers diverging from complex tracts like the arcuate fasciculus and the cingulum which innervate cortical and subcortical regions. This is observable in fig. 11. Furthermore, visual inspection of the *tract probability maps*, fig.s 13 and 14, shows that the mean bundles are consistent with probability maps obtained by manual selection of the bundles in Hua et al. (2008, Figure 3) and chemical staining on post-mortem brains (Bürgel et al., 2006). The high mean similarity with low dispersion on the cingulum, the cortico spinal tract and the inferior fronto-occipital fasciculus is consistent with histological studies (Bürgel et al., 2006). Additionally, the relatively lower and more disperse mean similarity on the arcuate and uncinate fasciculus is also consistent with previously cited histological studies.

While the method is fully automated, it depends on the parcellation used to produce the tracts. We noticed two cases where non-anatomically coherent low probability regions (yellow) are present: the left section of the fornix, fig. 13(c), and the left uncinate fasciculus, fig. 14(c). In the case of the left side of the fornix, fig. 13(c), even though the high probability region (red) is consistent with the previously cited anatomical studies, there is a low probability region corresponding to the optical nerve that can be seen in this map. Careful analysis of the clustered fibers generating this region shows that this artifact comes from the used tractography techniques; these fibers go all the way from the posterior column of the fornix to the frontal sections of the optical nerve. This is not anatomically correct and is the source of the high variability exhibited by this tract in the quantitative analysis, fig. 15. In the case of the left uncinate fasciculus, fig. 14(c), there is a high probability region (red) which is consistent with the previously cited anatomical studies but there is a low probability region corresponding to a section of the inferior-longitudinal fasciculus. This increases the variability of the clusters corresponding to the left uncinate fasciculus region, fig. 15. Analysis of the clustered fibers constituting this bundle and the query used to obtain the cluster corresponding to the uncinate fasciculus, table 1, shows that this artifact is produced by the white matter parcellation used to create the tracts. The absence of a parcel corresponding to the temporal pole in the atlas leads us to use the medial temporal gyrus for the corresponding query, shown in fig. 7 right side in light green. This allows fibers fully contained within the medial temporal gyrus to be incorporated in the uncinate fasciculus as a result of the corresponding query. In the future, we will investigate the use of alternate atlases or expert-corrected ROIs to remedy this problem.

6 Conclusion

We presented a mathematical framework to perform statistical operations on white matter fibers obtained from diffusion tensor MRI tractography. We showed that this framework spans an inner product space, and that this inner product space is appropriate for performing statistical operations among white matter bundles. We used this framework to build a clustering algorithm and a tract querying algorithm which allow automatic fiber bundle identification with the white matter queries as a sole parameter. Then, we applied these algorithms to 21 subjects registered to a publicly available atlas with a parcellation of the white matter gyri that we used as a priori anatomical information. Finally, we showed that the results of our clustering were consistent with studies carried by dissection macroscopical preparations and chemical staining. Thus, our framework has proved suitable to perform group studies of the cerebral white matter with minimal user interaction.

7 Acknowledgments

This work was partly supported by the INRIA ARC Diffusion MRI Program, the CORDI-S INRIA program and the Odyssée-EADS Grant 2118. Ragini Verma and Luke Bloy acknowledge support from the NIH grants R01MH079938 and T32-EB000814 respectively. The data was acquired as part of the NIH grant R01MH060722. Demian Wassermann wants to acknowledge Rebecca Wolpin for English language editing and Romain Veltz for helpful discussions.

A Gaussian Process framework for white matter fibers

Gaussian Processes (GP) are a rich mathematical framework that has been used in a wide variety of research fields. Examples of this being machine learning (MacKay, 1998; Rasmussen and Williams, 2006), spatial statistics (Stein, 1999) and modeling of observational data (Wahba, 1990). Furthermore, they are also related to kernel machines and radial basis functions (Wahba, 1990; Rasmussen and Williams, 2006). In this work we take advantage of the GP to produce a parametric representation of fiber bundles. Particularly, we take advantage of the capability to incorporate different types of hypotheses into the model. Moreover, this representation provides a simple way to linearly combine fibers into bundles and to measure similarity through an inner product operation which we develop in section A.3. Moreover, it provides a natural method to calculate *tract probability maps*, developed in section A.4.

A.1 Single fibers as Gaussian Processes

The parametric representation of a single fiber bundle is based on two hypotheses. Firstly, *smoothness*: due to the fact that fiber bundles in the brain do not have sharp angles (Basser et al., 2000), we consider that the least wiggled trajectory joining the sample points of a fiber

represents that fiber in a most probable manner. Secondly, *diffusion associated uncertainty*: the uncertainty of the position of each point on the bundle can be modeled using the water diffusion profile at that position (Basser and Pierpaoli, 1996).

Using these two hypotheses, we write $y_{\mathcal{F}}(\mathbf{p})$, the GP for the indicator function of fiber $\mathcal{F}$, $y(\mathbf{p})$ for clarity, as the combination of two other GPs, $z(\mathbf{p})$ and $\epsilon(\mathbf{p})$. The process $z(\mathbf{p})$ represents the *smoothness* of the trajectory in space, its parameters are inferred from the point sequence obtained through the tractography of a fiber: $\mathbf{f} = \{\mathbf{f}_1, \ldots, \mathbf{f}_{|\mathbf{f}|}\} \subset \mathbb{R}^3$. The process $\epsilon(\mathbf{p})$ adds a variability to the fiber at $\mathbf{p}$ using full *diffusion information* and it is inferred from the tensor field over the fiber $\Sigma(\mathbf{f}_1), \ldots, \Sigma(\mathbf{f}_{|\mathbf{f}|})$. We show now that $z(\mathbf{p})$ and $\epsilon(\mathbf{p})$ can be modelled as GPs by characterizing them through covariance functions.

A.1.1 Smoothness

We formulate a probabilistic prior for the probability of a function $z(\cdot)$ being the indicator function of a fiber. From the hypotheses that an indicator function with smaller curvature is a more probable representation of a given trajectory and that the original trajectory $\mathcal{F}$ is C^2. We express these hypotheses as the probabilistic prior

$$-\log(\mathbb{P}\{z(\cdot)\}) = \frac{1}{2}\int_{\Omega} |D^2 z(\mathbf{p})|^2 d\mathbf{p} + const,\ \Omega \subset \mathbb{R}^3 \tag{8}$$

where the linear operator D maps $z(\cdot)$ to its derivative. Then, we characterize $z(\cdot)$ from this prior. To begin with, we rewrite the prior over a finite sampling $\{\mathbf{p}_i\}_i \subset \Omega$. Thus, letting $[z(\mathbf{p}_i)]_i$ be the column vector of values of $z(\cdot)$ at each sampled point, we reformulate eq. 8 as

$$-\log(\mathbb{P}\{[z(\mathbf{p}_i)]_i\}) = \frac{1}{2}[z(\mathbf{p}_i)]_i^T\, D^2 (D^2)^T\, [z(\mathbf{p}_i)]_i + const \tag{9}$$

which is the p.d.f of a multivariate Gaussian with covariance matrix

$$S = (D^2(D^2)^T)^{-1} = (D^4)^{-1}.$$

Hence, the function $z(\cdot)$ is represented by the GP,

$$z(\mathbf{p}) \sim \mathcal{GP}(z^*(\mathbf{p}), s(\mathbf{p}, \mathbf{p}')).$$

In order to fully characterize the GP for $z(\mathbf{p})$, we need to find an explicit formulation for $s(\mathbf{p}, \mathbf{p}')$. We do this by first rewriting $S = (D^4)^{-1}$ as $D^4 S = I$. This reveals $s(\mathbf{p}, \mathbf{p}')$ to be the Green function of the fourth derivative operator (Williams and Fitzgibbon, 2007; Wahba, 1990)

$$\int_{\Omega} D^4(\mathbf{u}, \mathbf{w}) s(\mathbf{w}, \mathbf{v}) d\mathbf{w} = \delta(\mathbf{u} - \mathbf{v}). \tag{10}$$

Next, the solution for eq. 10 inside a sphere in $\mathbb{R}^3$ of radius R is

$$s(\mathbf{p}, \mathbf{p}') := \psi(\|\mathbf{p} - \mathbf{p}'\|), \psi(r) = \begin{cases} 2|r|^3 - 3Rr^2 + R^3 & r \leq R \\ 0 & r > R \end{cases}, \tag{11}$$

where the constants have been chosen such that $s(\mathbf{p}, \mathbf{p}')$ is a positive semi-definite symmetric function. Thus, we have fully characterized the probabilistic space of functions which describes the *blurred indicator function* for a smooth trajectory. Furthermore, the covariance function in eq. 11 is equivalent to a bounded thin-plate spline regularizer, which accurately models smooth spatial data (Wahba, 1990).

A.1.2 Diffusion associated uncertainty

We now represent the diffusion information of a point in the trajectory by a *blurring* GP at that point. The usual practice in GP literature is to model uncertainty at a given point by an isotropic blurring function (MacKay, 1998). In our case, we use anisotropical blurring at every given point sampled from fiber by means of a second covariance function based on Diffusion Tensor information. This covariance function is built with convolution kernels: Let $g(\cdot)$ be the GP of an isotropic blurring function, then its convolution with a kernel $k(\cdot)$,

$$\epsilon(\mathbf{p}) = \int k(\mathbf{w}; \mathbf{p}) g(\mathbf{w}) d\mathbf{w}$$

is a GP with covariance function (Paciorek and Schervish, 2006),

$$u(\mathbf{p}, \mathbf{p}') = \int k(\mathbf{w}; \mathbf{p}) k(\mathbf{w}; \mathbf{p}') d\mathbf{w}. \tag{12}$$

We express the *blurring* kernel $k(\cdot; \mathbf{p})$ as the probability of a particle going from $\mathbf{p}$ to $\mathbf{w}$ in a time τ in terms of the Diffusion Tensor $\Sigma(\mathbf{p})$ (Basser et al., 1994):

$$k(\mathbf{w}; \mathbf{p}) = \mathbb{P}\left\{\mathbf{w} | \mathbf{p}, \tau, \Sigma(\mathbf{p})\right\} = \\ = \frac{1}{\sqrt{(4\pi\tau)^3 |\Sigma(\mathbf{p})|}} e^{-\frac{1}{4\tau}(\mathbf{w}-\mathbf{p})^T \Sigma^{-1}(\mathbf{p})(\mathbf{w}-\mathbf{p})}.$$

Then, by performing the integral in eq. 12, we write the covariance function for the *anisotropical blurring process* as,

$$u(\mathbf{p}, \mathbf{p}') = \frac{1}{\sqrt{(4\pi\tau)^3 \left|\Sigma + \Sigma'\right|}} e^{-\frac{1}{4\tau}(\mathbf{p}-\mathbf{p}')^T \left(\Sigma+\Sigma'\right)^{-1}(\mathbf{p}-\mathbf{p}')},$$

where $\Sigma = \Sigma(\mathbf{p})$ and $\Sigma' = \Sigma(\mathbf{p}')$. The blurring at each point is represented by a zero mean GP,

$$\epsilon(\mathbf{p}) \sim \mathcal{G}P\left(0, u(\mathbf{p}, \mathbf{p}')\right).$$

A.1.3 Gaussian Process representation of a fiber

Finally, we write the GP formulation for the blurred indicator function of the fiber,

$$y(\mathbf{p}) \sim \mathcal{GP}\Big(y^*(\mathbf{p}) = z^*(\mathbf{p}); c(\mathbf{p}, \mathbf{p}') = s(\mathbf{p}, \mathbf{p}') + u(\mathbf{p}, \mathbf{p}')\Big). \tag{13}$$

Up to this point we have a Gaussian Process-based model for the white matter fibers. This model incorporates spatial and diffusion information. Moreover, within this model we can linearly combine fibers into bundles as show in eq. 2 and quantify similarity as shown in eq. 3. In the remainder of this section we show how to characterize the value of the indicator function $y(\mathbf{p})$ at any test point $\mathbf{p} \in \mathbb{R}^3$, how to effectively calculate the similarity between two fibers and finally, how to calculate the *tract probability map* for a bundle.

A.2 Calculating the indicator function value distribution for a test point

We want to calculate a p.d.f. for the value of $y(\cdot)$ at a test point $\mathbf{p}\in\mathbb{R}^3$. This is a simple operation as we have GP representation of the indicator function $y(\cdot)$ for a trajectory $\mathcal{F}$. The p.d.f. of $y(\mathbf{p})$, given the tractographed point sequence $\mathbf{f} = \{\mathbf{f}_1, \ldots, \mathbf{f}_{|\mathbf{f}|}\}\subset\mathbb{R}^3$ and its corresponding tensor field $\Sigma = \{\Sigma(\mathbf{f}_1), \ldots, \Sigma(\mathbf{f}_{|\mathbf{f}|})\}\subset SPD(3)$ are characterized as Gaussian-distributed random variable. More precisely, $y(\mathbf{p})$ is the Gaussian distribution,

$$(y(\mathbf{p})|\mathbf{f}, \Sigma, \mathbf{p}) \sim \mathcal{G}\left(y^*(\mathbf{p}), \sigma^2(\mathbf{p})\right) \tag{14}$$

due to marginalization properties of the GPs (MacKay, 1998). Setting $y(\mathbf{p})$ to take the constant value l when $\mathbf{p}$ is a point that belongs to the fiber trajectory $\mathcal{F}$, the mean and covariance functions can be calculated in the following way (MacKay, 1998):

$$y^*(\mathbf{p}) = S_{\mathbf{f}}(\mathbf{p})^T C_{\mathbf{ff}}\mathbf{1}l \qquad \sigma^2(\mathbf{p}) = s_{\mathbf{p}} - S_{\mathbf{f}}(\mathbf{p})^T C_{\mathbf{ff}}^{-1} S_{\mathbf{f}}(\mathbf{p}), \tag{15}$$

where $[S_{\mathbf{f}}(\mathbf{p})]_i = [s(\mathbf{f}_i, \mathbf{p})]_i$, $[C_{\mathbf{ff}}]_{ij} = [c(\mathbf{f}_i, \mathbf{f}_j)]_{ij}$ with $1 \leq i, j \leq |\mathbf{f}|$, $\mathbf{1}$ is the vector with all ones and $s_{\mathbf{p}} = s(\mathbf{p}, \mathbf{p})$; the functions $s(\cdot, \cdot)$ and $c(\cdot, \cdot)$ were defined in eq. 11 and eq. 13. This formulation is equivalent to "train" a Gaussian Process-based regression with values l at the sampled fiber points and 0 everywhere else. The parameter R is taken as the minimal distance between two consecutive points in $\mathbf{f}$. Examples of *blurred indicator functions* can be seen in figures 3 and 4.

A.3 Calculating the deterministic inner product between two fiber bundles

In order to produce an inner product space for fiber bundles, we must provide an inner product operation. We have previously introduce this operation in eq. 3. In this section we give a more formal definition and show how to calculate it efficiently.

We look for an inner product that quantifies fiber bundle overlapping in a *deterministic* manner. This simplifies the mathematical treatment of our inner product space, hence its applicability for large scale computations such as clustering of densely sampled full brain tractographies. In order to provide a *deterministic* inner product, we need to make a decision about the value of $y(\mathbf{p})$ defined in eq. 15. We do this by employing *decision theory* (De Groot, 2004): To make a decision about the value of $y(\mathbf{p})$ at $\mathbf{p}$, we use a point-like prediction, $y^+(\mathbf{p})$.

This prediction is taken in order to minimise the error in the squared norm induced by our inner product eq. 4,

$$\underset{y^+(\mathbf{p})}{\operatorname{argmin}} \int \left(y^+(\mathbf{p}) - y(\mathbf{p})\right)^2 \mathbb{P}\left\{y(\mathbf{p})|\mathbf{f},\mathbf{t},\mathbf{p}\right\} dy(\mathbf{p}) = y^*(\mathbf{p}). \tag{16}$$

Thus letting the mean value of $y(\mathbf{p})$, $y^*(\mathbf{p})$, be an appropriate estimator of the value of $y(\mathbf{p})$ at $\mathbf{p}$. Then, as $y^*(\mathbf{p})$ is square integrable due to its definition in eq. 15,

$$\langle \mathcal{F}, \mathcal{F}' \rangle := \int_{\mathbb{R}^3} y^*_{\mathcal{F}}(\mathbf{p}) y^*_{\mathcal{F}'}(\mathbf{p}) d\mathbf{p}$$

is an inner product (Schmidt, 1908). Furthermore, it can be easily computed by replacing eq. 15 in the previous formula:

$$\begin{aligned}\langle \mathcal{F}, \mathcal{F}' \rangle &:= \int_{\mathbb{R}^3} \left(S_{\mathbf{f}}(\mathbf{p})^T C_{\mathbf{ff}} \mathbf{1} l\right)^T \left(S_{\mathbf{f}'}(\mathbf{p})^T C_{\mathbf{f}'\mathbf{f}'} \mathbf{1} l\right) d\mathbf{p} \\ &= \int_{\mathbb{R}^3} \left(l \mathbf{1}^T C_{\mathbf{ff}}^T\right) S_{\mathbf{f}}(\mathbf{p}) S_{\mathbf{f}'}(\mathbf{p})^T \left(C_{\mathbf{f}'\mathbf{f}'} \mathbf{1} l\right) d\mathbf{p}.\end{aligned}$$

Moreover, we can simplify the previous expression using that $S_{\mathbf{f}}(\mathbf{p})$ and $S_{\mathbf{f}'}(\mathbf{p})$ are the only vectors depending on $\mathbf{p}$, see section A.2:

$$\langle \mathcal{F}, \mathcal{F}' \rangle := l \mathbf{1}^T C_{\mathbf{ff}}^T \left(\int_{\mathbb{R}^3} S_{\mathbf{f}}(\mathbf{p}) S_{\mathbf{f}'}(\mathbf{p})^T d\mathbf{p} \right) C_{\mathbf{f}'\mathbf{f}'} \mathbf{1} l$$

where

$$\left[\int_{\mathbb{R}^3} S_{\mathbf{f}}(\mathbf{p}) S_{\mathbf{f}'}(\mathbf{p})^T d\mathbf{p} \right]_{ij} = \int_{\mathbb{R}^3} s(\mathbf{f}_i, \mathbf{p}) s(\mathbf{f}'_j, \mathbf{p}) d\mathbf{p}$$

can be calculated analytically. Moreover, the similarity among two bundles of white matter fibers, $\mathcal{B} = \{\mathcal{F}_1, \ldots, \mathcal{F}_N\}$ and $\mathcal{B}' = \{\mathcal{F}'_1, \ldots, \mathcal{F}'_M\}$,

$$\langle \mathcal{B}, \mathcal{B}' \rangle = \left\langle \frac{1}{N} \sum_{i=1}^{N} y_{\mathcal{F}_i}(\mathbf{p}), \frac{1}{M} \sum_{j=1}^{M} y_{\mathcal{F}'_j}(\mathbf{p}) \right\rangle,$$

can be calculated in a simple manner by using the linearity and symmetry properties of the inner product operation. Then, the previous equation becomes

$$\langle \mathcal{B}, \mathcal{B}' \rangle = \frac{1}{NM} \sum_{i=1}^{N} \sum_{j=1}^{M} \left\langle y_{\mathcal{F}_i}(\mathbf{p}), y_{\mathcal{F}'_j}(\mathbf{p}) \right\rangle. \tag{17}$$

Examples of the deterministic inner product of two fibers and of the product of two mean indicator functions are shown in fig. 4. It can be noted that the inner product quantifies partial fiber overlapping.

A.4 Calculation of the tract probability map

Having characterized $y(\cdot)$ as a GP in section A.1, we can express the probability that a point $\mathbf{p}$ in $\mathbb{R}^3$ is contained in a bundle $\mathcal{F}$, this map is called the *tract probability map*. We do this by calculating the probability of $y(\mathbf{p}) = l$ or, equivalently the concentration of the random value $y(\mathbf{p})$ around l,

$$\mathbb{P}\{\mathbf{p}\in\mathcal{F}\} := \mathbb{P}\{y(\mathbf{p}) = l|\mathbf{f}, \Sigma, \mathbf{p}\} \propto \mathbb{E}[\theta(y(\mathbf{p}) - l)|\mathbf{f}, \mathbf{t}, \mathbf{p}]. \tag{18}$$

Then, we quantify concentration of $y(\mathbf{p})$ around l by the symmetric kernel $\theta : \mathbb{R}\rightarrow[0, 1]$. To ease the equations and the computation time, we take a Gaussian kernel, $\theta(x)$, with standard deviation h,

$$\theta(y(\mathbf{p}) - l) = \frac{1}{2\sqrt{\pi h}} \exp\left(-\left(\tfrac{y(\mathbf{p})-l}{h}\right)^2\right).$$

Then, we calculate eq. 18 as

$$\mathbb{E}[\theta(y(\mathbf{p}) - l)|\mathbf{f}, \mathbf{t}, \mathbf{p}] = \int \theta(y(\mathbf{p}) - l)\mathbb{P}\{y(\mathbf{p})|\mathbf{f}, \mathbf{t}, \mathbf{p}\}\, dy(\mathbf{p})$$

which leads to

$$\mathbb{P}\{\mathbf{p}\in\mathcal{F}\} \propto \mathbb{E}[\theta(y(\mathbf{p}) - l)|\mathbf{f}, \mathbf{t}, \mathbf{p}] = \frac{1}{2\sqrt{\pi(h^2 + \sigma^2(\mathbf{p}))}}. \tag{19}$$

where h is a bandwidth parameter and $\sigma^2(\mathbf{p})$ is defined in eq. 15. Then, the *tract probability map* for a bundle $\mathcal{F}$ on a domain Ω, is calculated by evaluating $\mathbb{P}\{\mathbf{p} \in \mathcal{F}\}$ at every point $\mathbf{p} \in \Omega$. In order to illustrate the probabilistic map for a bundle, color-coded surfaces and a probability map over an FA image for anatomical bundles are shown in fig. 5. These tract probabilistic maps are highly similar to the hand-obtained ones by Hua et al. (2008) on DTI images and Bürgel et al. (2006) by means of chemical staining. This two previous works have shown that this *tract probability maps* are an appropriate tool to perform statistics on white matter fiber bundles.

References

Basser, P., Mattiello, J., LeBihan, D., 1994. MR diffusion tensor spectroscopy and imaging. Biophysical Journal 66 (1), 259–267.

Basser, P., Pajevic, S., Pierpaoli, C., Duda, J., Aldroubi, A., 2000. In vivo fiber tractography using DT-MRI data. Magnetic Resonance in Medicine 44 (4), 625–632.

Basser, P., Pierpaoli, C., 1996. Microstructural and physiological features of tissues elucidated by quantitative-diffusion-tensor MRI. J. Mag Res B 111 (3), 209–219.

Batchelor, P. G., Calamante, F., Tournier, J.-D., Atkinson, D., Hill, D. L. G., Connelly, A., 2006. Quantification of the shape of fiber tracts. MRM 55 (4), 894–903.

Bürgel, U., Amunts, K., Hoemke, L., Mohlberg, H., Gilsbach, J. M., Zilles, K., 2006. White matter fiber tracts of the human brain: Three-dimensional mapping at microscopic resolution, topography and intersubject variability. NeuroImage 29 (4), 1092 – 1105.

Ciccarelli, O., Catani, M., Johansen-Berg, H., Clark, C., Thompson, A., 2008. Diffusion-based tractography in neurological disorders: concepts, applications, and future developments. Lancet Neurology 7 (8), 715–727.

Corouge, I., Fletcher, P. T., Joshi, S., Gouttard, S., Gerig, G., 2006. Fiber tract-oriented statistics for quantitative diffusion tensor mri analysis. MIA 10 (5), 786–798.

De Groot, M., 2004. Optimal Statistical Decisions. Classics Library. Wiley.

Descoteaux, M., Deriche, R., Knoesche, T., Anwander, A., 2009. Deterministic and probabilistic tractography based on complex fiber orientation distributions. Trans. in Med. Imag. 28 (2), 269–286.

Ding, Z., Gore, J., Anderson, A., 2003. Classification and quantification of neuronal fiber pathways using diffusion tensor MRI. MRM 49, 716–721.

Goodlett, C. B., Fletcher, P. T., Gilmore, J. H., Gerig, G., 2008. Group statistics of dti fiber bundles using spatial functions of tensor measures. In: MICCAI.

Hua, K., Zhang, J., Wakana, S., Jiang, H., Li, X., Reich, D. S., Calabresi, P. A., Pekar, J. J., van Zijl, P. C. M., Mori, S., 2008. Tract probability maps in stereotaxic spaces: Analyses of white matter anatomy and tract-specific quantification. NeuroImage 39 (1), 336–347.

Jain, A. K., Murty, M. N., Flynn, P., 1999. Data clustering: a review. ACM Computing Surveys (CSUR) 31 (3), 264–323.

Koch, M., Norris, D., Hund-Georgiadis, M., 2002. An Investigation of Functional and Anatomical Connectivity Using Magnetic Resonance Imaging. NeuroImage 16 (1), 241–250.

Kubicki, M., McCarley, R., Westin, C., Park, H., Maier, S., Kikinis, R., Jolesz, F., Shenton, M., 2007. A review of diffusion tensor imaging studies in schizophrenia. J Psych Res 41 (1-2), 15–30.

Lawes, I., Barrick, T., Murugam, V., Spierings, N., Evans, D., Song, M., Clark, C., 2008. Atlas-based segmentation of white matter tracts of the human brain using diffusion tensor tractography and comparison with classical dissection. NeuroImage 39 (1), 62–79.

Leemans, A., Sijbers, J., Backer, S. D., Vandervliet, E., Parizel, P., Jun 2006. Multiscale white matter fiber tract coregistration: a new feature-based approach to align diffusion tensor data. MRM 55 (6), 1414–1423.

Lenglet, C., Campbell, J., Descoteaux, M., Haro, G., Savadjiev, P., Wassermann, D., Anwander, A., Deriche, R., Pike, G., Sapiro, G., Siddiqi, K., Thompson, P., 2008. Mathematical methods for diffusion mri processing. Neuroimage.

MacKay, D. J. C., 1998. Introduction to gaussian processes. In: Neural Networks and Machine Learning. Vol. 168 of NATO ASI. Springer, pp. 133–165.

Maddah, M., Grimson, W. E. L., Warfield, S. K., Wells, W. M., 2008a. A unified framework for clustering and quantitative analysis of white matter fiber tracts. MIA, 191–202.

Maddah, M., Zollei, L., Grimson, W. E. L., Westin, C.-F., Wells, W. M., 2008b. A mathematical framework for incorporating anatomical knowledge in DT-MRI analysis. In: ISBI.

Mori, S., Crain, B. J., Chacko, V. P., Zijl, P. C. M. V., 1999. Three-dimensional tracking of axonal projections in the brain by magnetic resonance imaging. Ann Neur 45 (2), 265–269.

O'Donnell, L. J., May 2006. Cerebral white matter analysis using diffusion imaging. Ph.D. thesis, Massachusets Institute of Technology.

O'Donnell, L. J., Westin, C.-F., nov 2007. Automatic tractography segmentation using a high-dimensional white matter atlas. IEEE TMI 26 (11), 1562–1575.

O'Donnell, L. J., Westin, C.-F., Golby, A. J., 2007. Tract-based morphometry. In: MICCAI. pp. 161–168.

Oh, J. S., Song, I. C., Lee, J. S., Kang, H., Park, K. S., Kang, E., Lee, D. S., Jul 2007. Tractography-guided statistics (tgis) in diffusion tensor imaging for the detection of gender difference of fiber integrity in the midsagittal and parasagittal corpora callosa. Neuroimage 36 (3), 606–616.

Paciorek, C., Schervish, M., 2006. Spatial modelling using a new class of nonstationary covariance functions. Environmetrics (London, Ont.) 17 (5), 483.

Qazi, A. A., Radmanesh, A., O'Donnell, L., Kindlmann, G., Peled, S., Whalen, S., Westin, C.-F., Golby, A. J., 2008. Resolving crossings in the corticospinal tract by two-tensor streamline tractography: method and clinical assessment using fMRI. Neuroimage.

Rasmussen, C. E., Williams, C. K. I., 2006. Gaussian Processes for Machine Learning. The MIT Press.

Savadjiev, P., Campbell, J., Descoteaux, M., Deriche, R., Pike, G., Siddiqi, K., 2008. Labeling of ambiguous subvoxel fibre bundle configurations in high angular resolution diffusion MRI. NeuroImage 41 (1), 58–68.

Schmidt, E., 1908. Über die auflösung linearer gleichungen mit unendlich vielen unbekannten. Rend. Circ. Mat. Palermo 25, 63–77.

Stein, M. L., 1999. Statistical Interpolation of Spatial Data: Some Theory for Kriging. Springer.

Veltkamp, R., 2001. Shape matching: Similarity measures and algorithms. In: Shape Modeling and Applications, SMI. pp. 188–197.

Wahba, G., 1990. Spline Models for Observational Data. Soc Ind Math.

Wakana, S., Caprihan, A., Panzenboeck, M. M., Fallon, J. H., Perry, M., Gollub, R. L., Hua, K., Zhang, J., Jiang, H., Dubey, P., Blitz, A., van Zijl, P., Mori, S., 2007. Reproducibility of quantitative tractography methods applied to cerebral white matter. NeuroImage 36 (3), 630 – 644.

Wakana, S., Jiang, H., Nagae-Poetscher, L., van Zijl, P., Mori, S., 2004. Fiber Tract–based Atlas of Human White Matter Anatomy. Radiology 230, 77.

Wassermann, D., Deriche, R., 2008. Simultaneous manifold learning and clustering: Grouping white matter fiber tracts using a volumetric white matter atlas. In: MICCAI Workshops.

Wiegell, M., Larsson, H., Wedeen, V., 2000. Fiber Crossing in Human Brain Depicted with Diffusion Tensor MR Imaging 1. Radiology 217 (3), 897–903.

Williams, O., Fitzgibbon, A., 2007. Gaussian process implicit surfaces. In: Gaussian Proc. in Practice.

Yang, J., Shen, D., Davatzikos, C., Verma, R., 2008. Diffusion Tensor Image Registration Using Tensor Geometry and Orientation Features. In: MICCAI.

Unité de recherche INRIA Sophia Antipolis
2004, route des Lucioles - BP 93 - 06902 Sophia Antipolis Cedex (France)

Unité de recherche INRIA Futurs : Parc Club Orsay Université - ZAC des Vignes
4, rue Jacques Monod - 91893 ORSAY Cedex (France)
Unité de recherche INRIA Lorraine : LORIA, Technopôle de Nancy-Brabois - Campus scientifique
615, rue du Jardin Botanique - BP 101 - 54602 Villers-lès-Nancy Cedex (France)
Unité de recherche INRIA Rennes : IRISA, Campus universitaire de Beaulieu - 35042 Rennes Cedex (France)
Unité de recherche INRIA Rhône-Alpes : 655, avenue de l'Europe - 38334 Montbonnot Saint-Ismier (France)
Unité de recherche INRIA Rocquencourt : Domaine de Voluceau - Rocquencourt - BP 105 - 78153 Le Chesnay Cedex (France)

Éditeur
INRIA - Domaine de Voluceau - Rocquencourt, BP 105 - 78153 Le Chesnay Cedex (France)
http://www.inria.fr
ISSN 0249-6399